Home Wi-Fi Impairments

Márquez-Israel Salinas

Home Wi-Fi Impairments

Israel Márquez Salinas
UPMC Sorbonne Universités/INRIA

Supervisors: Renata Teixeira, Timur Friedman
Advisors: Sara Ayoubi, Francesco Bronzino
Inria, UPMC

ABSTRACT

The preferred method to access Internet from home is Wi-Fi. Unfortunately poorly placed Wi-Fi access point can experience Wi-Fi impairments such as interference or congestion, leading to degraded Internet performance. Identifying these impairments can be challenging, even for wireless experts. To approach this challenge we develop a tool to identify home Wi-Fi impairments. In our work we conduct experiments triggering wireless and non-wireless issues in a testbed. The two methods we work with are active probing and passive wireless metrics collection from wireless AP and wireless client. The wireless metrics we collect include but are not limited to, RSSI, PHY Rate, Noise, etc. With these metrics we get a sense of the status of home Wi-Fi and correlate it with our active probing results. Finally, to identify a wireless impairment we run our dataset through supervised learning algorithms. We obtain the best results with random forest algorithm. Random forest is well known for its precision to classify events based on a specific set of features. We close our paper by presenting the results of home Wi-Fi impairment detection by modeling it as a classification problem.

1. INTRODUCTION

The most common way to access Internet from home is Wi-Fi. The variety of services and devices using the home Wi-Fi to access Internet is vast. It is common today for a home user to stream a movie on his laptop while connected to the home Wi-Fi. In some cases, the movie streaming is degraded, which is frustrating for the users. One of the potential causes of poor streaming experience is the home Wi-Fi. In fact, previous work [11] has identified home Wi-Fi as the bottleneck along the end-to-end path. The cause of poor home Wi-Fi experience can be varied [5], channel congestion, poor client or AP placement and interference are the most common causes. Other work [8] has analyzed the impact of Home Wi-Fi on the latency in a network path. They have identified that Wi-Fi latency can contribute up to 60% of the overall round trip time along the end-to-end path. ISPs are often held responsible for poor Internet experience [1]. Home users, in the search for a solution can switch between ISPs or content providers even though the problem is within the home. In this research paper we develop a tool to identify home Wi-Fi impairments. We describe the initial stages of this tool in this paper.

Identifying where the root cause is within the home Wi-Fi is challenging due to multiple factors. First, wireless nature is volatile as it uses an open and shared medium, shared among Wi-Fi and non-Wi-Fi devices. Second, it is required to have a vantage point within the home. This vantage point should be common across home deployments and capable of collecting Wi-Fi metrics to assist on the identification of Wi-Fi impairments. Most research work has mainly implemented passive techniques to identify where the potential cause for a degraded service is located [3], [2], [7]. A couple others have relied upon active techniques [4], [12]. Depending on the type of measurement technique, it is required to address different considerations. Passive techniques face the challenge of requiring access to the AP to collect the metrics. Making changes to the AP to collect metrics is another challenge as most APs are not open to be customized. With active techniques the complication is the potential overhead caused by the measurement tool. In other words, with active techniques the network can experience disruption.

Our tool implements both techniques to take strong points of both and leverage the weakness with each other's strong points. In this initial phase we are using both techniques to identify the relationship between metrics passively collected and active probing results. We believe that mainly relying on active probing to identify home Wi-Fi impairments can be a breakthrough in the development of tools to be widely deployed in the wild. Further description of these techniques is covered in Section 2. Related work associated to home Wi-Fi study will be covered in Section 5. The instrumentation details of our tool are developed in Section 3. The mechanisms and techniques to evaluate the method used to identify home Wi-Fi impairments is explained in Section 4. Finally findings of our work are consolidated in Section 6.

2. WI-FI MONITORING

Active and passive techniques have their advantages and disadvantages. In the following, we outline the main characteristics of each one of them. Each of the techniques will be best-suited depending on the goal and context of the experiment.

2.1 Active

Active measurement is a technique in which traffic is injected in the network to get a sense of the network status. The injected packets are called probes. For our work we use active measurements to obtain metrics on bandwidth, Round-Trip Time (RTT) and packet loss. In the Wi-Fi context, bandwidth active measurements can help to identify where the bottleneck is happening. We have also used active bandwidth measurement tools to generate traffic in our experimental setup to resemble real-case scenarios. High RTT can help to identify if congestion is happening in the home Wi-Fi. In a similar way, packet loss can denote interference as frames are destroyed in the Wi-Fi link. While working with active measurements it is important to pay attention to the probes size and probing rate. Large probe sizes and aggressive probing rate can cause overhead. Overhead does not only disrupts user traffic but can also lead to biased measurements. In the following bullet points we outline the strengths and weaknesses of active measurements that we consider relevant for our work. We also include the ones we work with for this paper.

Strengths

- Full ownership of the network is not required.
- They do not require large space to store data collected as generally, probe packets are small.
- Privacy concerns are minimal as probe packets used to measure are made of random data which has no sensitive information.
- Useful to get the state of the network on-demand.

Weaknesses

- Overhead might occur if probe size and probing rate are chosen without due diligence of network conditions.
- Biased results can be obtained if probes, either size or rate, cause overhead in the network.

Under the scope of active measurement techniques, the following are the metrics to be actively collected for our work.

- **Round Trip Time**
  - For our goal, RTT can helps us identify if we are experiencing attenuation and interference in the home Wi-Fi. High RTT values can give a sense of latency in the home Wi-Fi which is potentially correlated to attenuation. Packet loss in the other hand, will point to interference related impairments as frames are being destroyed, causing the loss of these frames.
- **Throughput**
  - In Wi-Fi, throughput active measurement can assist to identify if congestion is happening in the Wi-Fi link. For example, if the AP reports a strong signal to the wireless client and minimal losses but the throughput is low, it is likely that the AP is experiencing congestion.

2.2 Passive

Passive measurement techniques rely on a "sit and listen" approach. The instrument conducting passive measurements in the network sits in a specific location along the path and records the metrics of interest. The monitor can be a component of the network itself, for example a router. It can also be device devoted to measure, such as a wireless sniffer. An important difference between active and passive techniques is that the latter do not trigger probes. Overhead due to probe packets is not present in passive measurements. However, computational and storage resources in the passive measuring device can be important factors to consider. The device might require to have enough space to store the data being collected. In a similar way, the computational power of the device can be required to be high depending on the speed of the link being measured. A Gigabit link in a core router will handle significantly more data than an 100Mbps Ethernet link in an access switch. Outlined in the following list a high level summary of the key strength and weaknesses of passive measurement techniques for our work purposes. We also outline the ones we use for our work.

Strength

- No extra traffic is generated to collect metrics, risk of causing overhead is minimized.

Weaknesses

- Large storage capacity can be required to store collected data. Not all measuring devices have large storage capacity, i.e. Access Points.
- Access to equipment working as passive measurement device is required. This is not possible for most users at multiple devices along an Internet path.
- High computational power on the measuring device can be required depending on the link being monitored and data granularity pursued. Not all devices can provide high computational power, i.e. Access Points.

- **RSSI - Received Signal Strength Indicator**
  - In our experiments we collect the RSSI from the AP and the wireless client. The RSSI help us to identify if there is attenuation happening in the link. A low RSSI denotes attenuation in the wireless link.
  - Low RSSI can be caused by poor AP placement due to large distance between wireless client and AP or, obstacles between both.
- **PHY Tx Rate**
  - The PHY Tx Rate at the wireless nodes can be an indicator of poor Wi-Fi link quality. A low PHY Tx Rate can help to diagnose a congestion, attenuation or interference Wi-Fi impairment. In collaboration with other metrics the scope of the impairment can be narrowed down.
  - For example, if the RSSI is strong, meaning there is no attenuation; loss rate is minimal, meaning interference is not present, but Tx PHY Rate is low; the impairment scope can be narrowed down to congestion.
- **Noise**
  - Noise measurements assist know if environment where the wireless client or the AP is placed is suitable for Wi-Fi. For example, if the noise level at a particular wireless client is high, we expect that node to be the only one with Wi-Fi degraded quality. In the other hand, if the AP is the one sensing high noise levels, we can expect all the clients connected to that AP to experience degraded Wi-Fi.
  - Noise can be caused by devices which "do not speak Wi-Fi language" such as microwave ovens, cordless phones and similar. Noise can help to distinguish between congestion and interference. Unlike congestion, interference is driven by non-WiFi sources.
- **Throughput - Driver Logs**
  - The throughput from the driver logs assist us to sense a Wi-Fi impairment. Low throughput can be an indicator of congestion, attenuation or interference.
  - In a similar way as for actively measuring throughput, collaborating with other metrics can narrow down the potential Wi-Fi impairment.
  - Additionally, passively measuring bandwidth helps to validate we obtain similar values obtained from active measurement techniques.
- **Frame Delivery Ratio**
  - Frame Delivery Ratio depicts the ratio between packets successfully received and total packets sent. The FDR metric can assist to get a sense of link quality. Low FDR indicates poor link quality.
  - Poor link quality can be caused due to congestion, attenuation or interference. In a similar way as with other metrics in our work, we collaborate with other metrics to narrow down the potential Wi-Fi impairment being experienced.

The metrics described before have been collected from different devices in our setup. Previous works have identified that even with similar wireless conditions devices can experience different throughput and bitrates [10], therefore we use different vantage points. Passive metrics have mostly been collected at the wireless client and AP. We extract these metrics from driver logs and derived statistic from them. Additionally we setup a wireless sniffer to get wireless captures. We use the wireless captures to validate the values we get from the logs at wireless client and AP. In the case of active metrics, we collect them from a wired client. The wired client works as the device in which we target to deploy our tool. From the wired client we trigger the active probing tool to collect RTT and throughput. The RTT measurements are collected using a custom ping-like tool. The active throughput is collected with *iPerf*. In section 3 we share instrumentation details on the tools we use to obtain the metrics we work with.

2.3 Finding the probing rate

Finding the probing rate is important when working with active measurements. A high rate can cause overhead, whereas a low rate can fail to capture the status of the network. To approach this challenge we conducted experiments in our office lab. Our experiments consisted in sending a series of ping trains which included multiple pings inside each train. We ran the tests with different train inter-spacing values and with different amount of pings inside the trains. The first finding from our experiments was a delay in RTT due to power save mode in devices. The power save mode sends the NIC to sleep. We refer to this delay as the "sleeping NIC". We found that when the inter-train spacing is smaller or equal to 100 msec the power save mode delay is not present. Based on this finding we set our lower bound for inter-train spacing to 100 msec. We set our upper bound to 1000 msec as the RTT within a single-hop home Wi-Fi network without significant cross-traffic is expected to be only a few milliseconds. This observation is remarked in the work of Sundaresan et al. [11].

The second relevant finding is associated to the RTT value of each ping within a train. We found that even with inter-train spacing values above 100 msec it is possible to overcome sleeping NIC delay by considering the RTT value of the 3rd or greater ping within a train. We noticed that the RTT value for ping greater or equal to the 3rd ping in a train depicted similar RTT values as when the sleeping NIC delay is not present. After these observations we defined our baseline to be 100 msec inter-train spacing and 3 pings per train. Figure 1 illustrates the values for the average round trip time of three pings in a train. The inter-ping spacing is equally distributed among the number of pings in a train and depends on the inter-train spacing. For example, the inter-ping spacing value for 3 pings in a 100 msec inter-train series is 100 msec/3 or 33.33 msec.

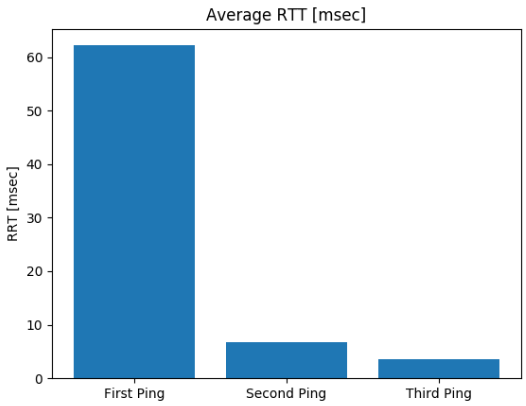


Figure 1: Average RTT for Three Ping Series

With this exercise we defined our baseline, we implemented similarity tests between our baseline results and samples derived from the baseline. We refer to our baseline as aggressive probing. To keep the samples to follow the same distribution as our baseline we implemented a Poisson process to generate the inter-train space intervals. In other words, randomly sampling from a Poisson process will result in another Poisson-distributed process [9]. This feature has been included in our GoPing tool. We sampled our baseline to obtain from 10% to 90 % of our original data points. We implemented Bernoulli random sampling to extract our samples. Finally, we ran Two-sample Kolmogorov-Smirnov tests between our baseline and samples. From the results we noticed that the sample which delivers a similar ECDF to our baseline is the one that keeps 50% of the original baseline data points. Figure 2 illustrates both ECDFs.

In figure 2 we can see the overlap between the sample keeping 50% of original data points and the original baseline derived from aggressive probing. This was the sample in which the overlap occurred, based on this result we set our active probing rate. As the RTT ECDF

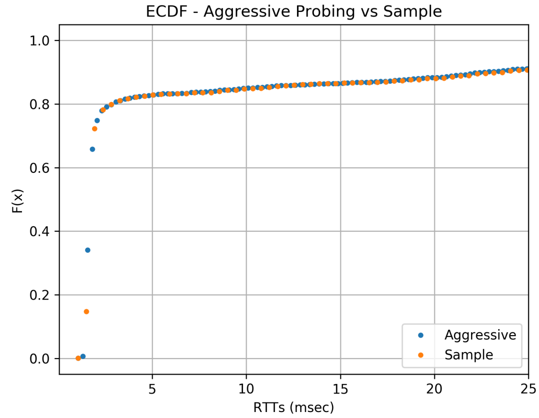


Figure 2: ECDF - Aggressive Probing vs Sample

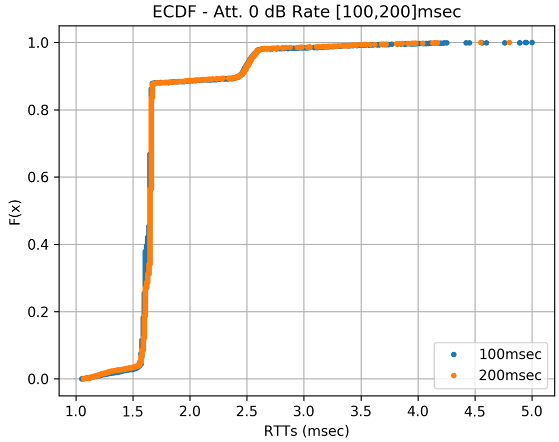
of the sample with half of the original data is similar to the original baseline, we set our probing rate to be 200 msec.

After finding the active probing rate as described in Section 2.3 we tested it in the testbed where we ran our experiments. Further description on our labs setup is covered in Section 3.1. The tests consisted in sending as many batches as possible for 10 min at 100 and 200 msec probing rates. Additionally we varied the attenuation with values of 0, 15 and 30 dBm. Table 1 summarizes the values used for the test. The test sessions took place in the 2.4 GHz band using 802.11n WLAN with no authentication. We ran each experiment session 5 times, in total we obtained 30 samples.

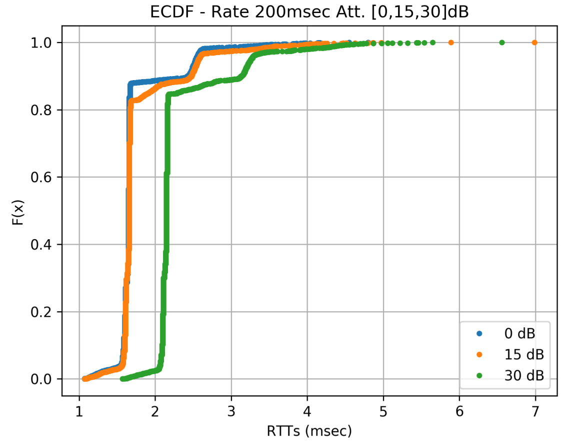
Attenuation	Probing Rate
0 dBm	100msec
0 dBm	200msec
15 dBm	100msec
15 dBm	200msec
30 dBm	100msec
30 dBm	200msec

Table 1: Attenuation and Probing Rate Validation Values

We compared the RTT ECDF of both to check similarity between probing rates. The expectation was for curves to be similar to each other. As expected, figure 3a illustrates the similarity between both RTT ECDF probing rates. The next expected behavior was for RTT to increase as the attenuation values increases. Figure 3b help us to validate the expected behavior. As we increase attenuation, RTT increases.



(a) Att. 0 dBm - Rate 100,200 msec



(b) Rate 200 msec - Att. [0,15,30] dB

Figure 3: RTT ECDFs for Attenuation and Probing Rate in Testbed

3. WIRELESS BOTTLENECK DETECTOR

We strive to detect Home Wi-Fi impairments, to achieve our goal we model our detector as a classification problem. To generate the data to train our model we ran experiments and trigger Wi-Fi impairments during the experiments. From the experiments we collect the metrics described in Section 2 and create datasets derived from them. Finally we consolidate our datasets and feed them into a supervised learning tool we used for classification.

3.1 Labs Setup

In order to run our experiment we work with two labs. A small In-office lab and a testbed to run experiments mostly associated to wireless technologies.

In-Office Lab

The In-office lab was mainly used to work on finding the probing rate to be used in the experiments at the testbed we worked with. The setup at our office is primarily composed by the following elements.

1. Raspberry Pi 3 running Raspbian GNU/Linux 8 (Jessie)
2. Wireless Access Point TP-Link AC1750
3. Dell Inspiron Laptop running Ubuntu 16.04.4 LTS (Xenial Xerus)

The wireless card driver on the Dell Laptop supports 802.11 a/b/g/n/ac. The driver is *iwlwifi* version 4.4.0-130-generic and firmware=17.948900127.0. At the In-Office lab we setup our deployment as illustrated in figure 4. The Pi was the device from which we send the pings towards the Laptop. As illustrated in figure 4 the Pi and access point are connected via Ethernet. The laptop is connected with the AP via 802.11n. During the tests at the In-lab office we switched between 2.4 and 5.0 GHz band depending on the goal of the experiment. The In-Office lab played a key role to test the features we included in our GoPing tool prior to running experiments at the second lab.

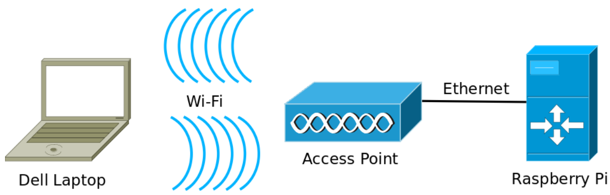


Figure 4: In-Office Lab Deployment

Orbit Lab

The second lab we work with is Orbit Lab [6]. Orbit lab is a large testbed in which different wireless technologies can be tested. One of these wireless technologies is 802.11. Within Orbit Lab we work with Sandbox 4 (SB4) which includes features to vary the attenuation between the nodes in the Sandbox. The main components of SB4 we work with are the followings.

1. SB4 has 9 nodes, each one of them runs Ubuntu 12.04
2. Attenuation Controller which makes possible to vary the attenuation between the nodes.

Each of the nodes has an Atheros Wireless card, the models are Atheros 5K and 9K. The nodes we work with have Atheros chipsets which allow us to collect detailed Wi-Fi logs. The links between the nodes in SB4 can be set to attenuation values between 0-30 dBm from the attenuation controller. The topology of SB4 depends on the attenuation values for each link. For example, a full-mesh topology is achieved when the attenuation value for all the links is set to 0 dBm. Thanks to the attenuation controller we can setup experiments in which changes on the RSSI are visible. The RSSI, as descried in section 2, is a metric which can help to identify attenuation impairments. The main deployment we use in Orbit is similar to the deployment we have at the In-office lab. The node working as AP was setup using the *hostapd* utility. The WLAN settings for the AP are summarized in table 2.

Setting	Value
802.11 Protocol	802.11n
Channel Bonding	No
Band	2.4 GHz
Security	Open

Table 2: WLAN Settings at AP Node

3.2 Experiments Setup

Once our labs were setup we ran our experiments. During our experiments we collect active and passive metrics using a diverse set of tools. Most of the tools we work with are out-of-the-box tools, such as iPerf, tc and tcpdump. The active measurement tool we use to collect RTTs is a custom implementation of Ping in GoLang. We refer to this tool as *GoPing*. We have customized GoPing to send ping trains and batches. To trigger Wi-Fi and non-Wi-Fi impairments we setup three main scenarios in Orbit SB4 testbed. Two of them, attenuation and congestion, are Wi-Fi-related. The third one is an access link impairment. Across the three scenarios we let running the collection of active and passive measurements. Our experiment sessions last 10 min. Passive wireless metrics logs are collected every 10 secs at the AP and Wireless client. We also setup a wireless a sniffer to obtain Over-the-Air packet captures. From the wired client we send pings towards the AP and log the stats from GoPing. For the throughput measurement with iPerf, we setup iPerf server at the wireless node and the iPerf client at the wired node. iPerf was setup in TCP mode with 4 parallel TCP streams. We chose TCP as we expect it to be the transport protocol most commonly used by services at home Wi-Fi networks. We work with 4 parallel streams as we expect the number of TCP streams in a home to be less than 5 TCP streams.

Attenuation

To trigger attenuation impairments in the testbed we vary the attenuation at the link between the wireless client and the access point. As mentioned before the setup we use at SB4 in Orbit is similar to our In-Office lab which is illustrated in figure 4. The additional component to this setup is the node working as a wireless sniffer. The attenuation impairment happens in the 3^{rd} and 9^{th} minute of the 10 min session. The experiment is designed this way to set a comparison between impairment-free and impairment conditions. This experiment design also assists on the evaluation of the classification tool we work with. We setup 5 scenarios with an increase of 3 dBm per impairment interval. Table 3 breakdowns the scenarios and the attenuation levels for each impairment interval. We ran each experiment 5 times.

Scenario	Attenuation Value [dBm]	
	1st Interval	2nd Interval
1	3	6
2	9	12
3	15	18
4	21	24
5	27	30

Table 3: Attenuation Scenarios and Values

Congestion

To trigger congestion in our testbed, we connect more wireless clients to the same AP our main wireless client is connecting to. At the 4^{th} and 8^{th} minute we connect an additional wireless clients to the AP. The additional wireless clients send UDP traffic to the AP using iPerf, hence the AP is running an iPerf server instance. We increase by one the wireless nodes connecting to the AP per interval. Table 4 consolidates the scenarios and number of wireless nodes connecting to the AP per interval.

Scenario	Connected Wireless Nodes	
	1st Interval	2nd Interval
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5

Table 4: Congestion Scenarios and Values

Access Link Limiting

The third scenario triggers a non-Wi-Fi impairment. The experiment consists in limiting the access link capacity at the wired node. We achieve limiting at the wired node by using *tc*, a traffic shaper utility. In a similar way to the attenuation scenario, we trigger the impairment conditions at the 3rd and 9th minute of the 10 min experiment window. Table 5 details the scenarios and values for the Access Link Limiting scenario.

Scenario	Throughput [Mbps]	
	1st Interval	2nd Interval
1	100	90
2	80	70
3	60	50
4	40	30
5	20	10

Table 5: Access Link Limiting Scenarios and Values

3.3 RSSI in the wild

With regards to the attenuation scenario we ran a survey to find the common RSSI value in home and office environments. We asked the colleagues at our office to run a script which collects Wi-Fi metrics on their laptops at different times during their stay at home or office. From the 760 samples collected we found that the most common RSSI value ranges between -60 and -65 dBm. Figure 5 illustrates the RSSI histogram of our survey.

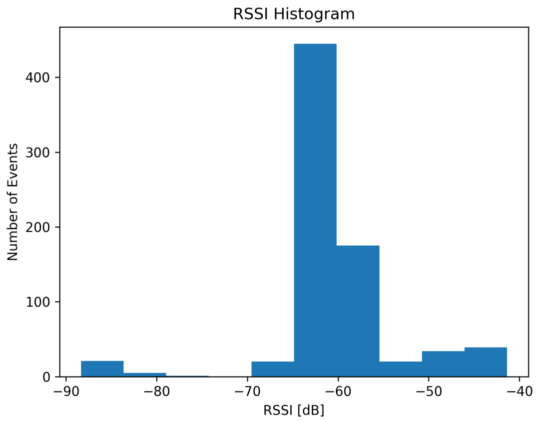


Figure 5: RSSI Survey Values Histogram

The main goal of this exercise is to validate attenuation values to be used in our experiments. In Orbit SB4 testbed, attenuation values of 0, 3 and 6 [dBm] lead to RSSI values between -60 and -65 dBm, which is the range obtained from our survey. These values are covered in our attenuation experiments.

3.4 Classification tool

After completing our experiments we moved on to the classification phase. To classify our events we use *Weka*. Weka is a software which has different supervised learning algorithms. We work with three algorithms available in Weka, J48, AdaboostM1 and random forest. To train Weka we feed it with the dataset derived from our experiments. Our experiments are designed in a way in which each impairment happens during a specific interval within the 10 min experiment window. Therefore we know at which interval does the impairment occurs. In other words, each experiment session consists in 10 intervals, each one 60 sec long. Depending on the scenario; congestion, attenuation or access link, we trigger the impairment at a specific interval as described in Section 3.2.

4. EVALUATION METHOD

To evaluate the classification generated by Weka we create two datasets, binary and multiclass. We create these datasets to feed them into Weka and evaluate the results with different algorithms. In the binary dataset we work with only two labels. The goal is to identify if there is a Wi-Fi impairment or not. The labeling schema for the binary dataset is described in table 7.

Label	Issue Type
0	No Issue at all
1	Attenuation
0	Access-Link Limiting
1	Congestion

Table 7: Binary dataset labels

As represented in table 7, even though "access link limiting" is a network impairment it shares the same label as "no issue at all". They both share the same label as access link limiting nature is non-Wi-Fi. The second type of dataset is multiclass. The multiclass dataset has the goal to classify in more detail the impairment, to distinguish if it is congestion, access link limiting, attenuation or none of them. In multiclass dataset we assign a different label to each impairment. The labeling for the multiclass dataset is described in table 8.

Label	Issue Type
0	No Issue at all
1	Attenuation
2	Access-Link Limiting
3	Congestion

Table 8: Multiclass dataset labels

For both datasets we ran the algorithms J48, AdaBoostM1 and random forest. We used the default 10 fold cross-validation in Weka for these algorithms. The results obtained from the binary dataset are outlined in table 6. The best results are obtained with the "random forest" algorithm. The next step was to feed Weka with the multiclass dataset. As mentioned before, with the multiclass dataset the goal is to classify the impairment in more detail. The results are summarized in table 6. Once again the best results are obtained with the random forest algorithm.

5. RELATED WORK

The challenge to identify impairments in the home Wi-Fi has been approached before. To address this challenge the research community has relied on two measurement techniques, active and passive. While most of previous works have opted for passive techniques [3], [2], [7]; others, have worked with active ones [4], [12]. The work of Da Hora et al [2] chose passive techniques. In their work they develop a method to detect poor QoE derived from Wi-Fi quality metrics. In their research context, active techniques have not been used to prevent user traffic disruption and battery drain of devices under study. Within the context of passive metrics, they excluded per packet analysis as it can result in overhead during high network utilization periods. Our efforts learn from their work on the meaningful Wi-Fi quality metrics to be passively collected. On the active probing side, we approach it in a different way as we implement an active probing rate which we believe will not cause user traffic disruption. In the work of Ashish et al [7], they present a metric called *Witt* which can help to get a sense of the home Wi-Fi experience from the AP perspective. Their implementation relies on passive measurements. Their AP, which is a custom AP, requires to be customized in hardware and software to interact with their infrastructure. From their work we takeaway the strong point of using the AP as a vantage point to get a sense of the home Wi-Fi experience. However, at this point we discard AP customization even though it increases the granularity and number of metrics collected. Our work pursues a software and hardware agnostic tool to facilitate its wide deployment in the wild and we believe AP customization can hinder this goal. On the active measurements techniques side, Kanuparthy et al [4] have implemented user-level probing. In their work they describe the ability to identify wireless pathologies derived from a metric proposed by them called wireless access delay. Their metric reflects the delays a packet faces while going through a 802.11 link. Their setup is almost agnostic as they require to deploy a wired device in the home Wi-Fi for their metric to be collected. Our work is similar to them as we also implement user-level probing. We use a ping-like tool to pursue software and hardware agnosticism. We approach the requirement of having a wired device deployed in the home Wi-Fi by leveraging with an existing project which facilitates the use of a wired device [3]. Under the same light of active measurement techniques, the work of Syrigos et al [12] defines a set of metrics to characterize Wi-Fi pathologies. The metrics proposed are derived from statistic available in most wireless devices. From their work we learn active metrics available in wireless equipment from which Wi-Fi impairments pathologies can be characterized. For the development of this paper we have learned from research mention along this section and continue to extend our knowledge on novel frameworks to approach the challenge of home Wi-Fi impairment detection.

6. CONCLUSION

In this initial stage the results obtained from random forest classification are promising nevertheless they can be considered "too good to be true". We consider that it is required for us to include more variability to the experiments we are working with and the size of the datasets to increase the robustness of our results. As mentioned in the opening of this paper, the goal is to include the findings of this initial stage to leverage the wide deployment of a home Wi-Fi impairments detector in the wild. We forecast leverage its deployment with existing projects already deployed in the wild. We are aware there is room for improvement in our work and we strive to further develop our methods to identify home Wi-Fi impairments.

Attribute-Algorithm	Dataset	Correctly Classified	Incorrectly Classified	RMSE	Precision	Recall
J.48	Binary	99.47%	0.53%	0.06	0.99	0.99
	Multiclass	98.66%	1.33%	0.07	0.98	0.98
Adaboost	Binary	99.73%	0.27%	0.05	0.99	0.99
	Multiclass	85.73%	14.26%	0.20	0.99	0.85
Random Forest	Binary	100.00%	0.00%	0.03	1.00	1.00
	Multiclass	99.86%	0.13%	0.04	0.99	0.99

Table 6: Classification Results

7. ACKNOWLEDGEMENTS

I would like to thank the supervisors and advisors who guided me on this work. I also thank the colleagues at Inria, LINCS and UPMC who contributed to this paper. This work was supported by the French ANR Project no. ANR-15-CE25- 0013-01 (BottleNet), by the Inria Project Lab BetterNet, and by a Google Faculty Award.

8. REFERENCES

[1] Diego Da Hora, Karel Van Doorselaer, Koen Van Oost, and Renata Teixeira. *Predicting the effect of home Wi-Fi quality on QoE*. PhD thesis, INRIA; Technicolor; Telecom ParisTech, 2018.

[2] Diego Neves da Hora, Karel Van Doorselaer, Koen Van Oost, Renata Teixeira, and Christophe Diot. Passive wi-fi link capacity estimation on commodity access points. In *Traffic Monitoring and Analysis Workshop (TMA) 2016*, 2016.

[3] Diana Joumblatt, Renata Teixeira, Jaideep Chandrashekar, and Nina Taft. Hostview: Annotating end-host performance measurements with user feedback. *ACM SIGMETRICS Performance Evaluation Review*, 38(3):43–48, 2011.

[4] Partha Kanuparthy, Constantine Dovrolis, Konstantina Papagiannaki, Srinivasan Seshan, and Peter Steenkiste. Can user-level probing detect and diagnose common home-wlan pathologies. *ACM SIGCOMM Computer Communication Review*, 42(1):7–15, 2012.

[5] Kyung-Hwa Kim, Hyunwoo Nam, and Henning Schulzrinne. Wislow: A wi-fi network performance troubleshooting tool for end users. In *INFOCOM*, pages 862–870, 2014.

[6] Maximilian Ott, Ivan Seskar, Robert Siraccusa, and Manpreet Singh. Orbit testbed software architecture: Supporting experiments as a service. In *Testbeds and Research Infrastructures for the Development of Networks and Communities, 2005. Tridentcom 2005. First International Conference on*, pages 136–145. IEEE, 2005.

[7] Ashish Patro, Srinivas Govindan, and Suman Banerjee. Observing home wireless experience through wifi aps. In *Proceedings of the 19th annual international conference on Mobile computing & networking*, pages 339–350. ACM, 2013.

[8] Changhua Pei, Youjian Zhao, Guo Chen, Ruming Tang, Yuan Meng, Minghua Ma, Ken Ling, and Dan Pei. Wifi can be the weakest link of round trip network latency in the wild. In *INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications, IEEE*, pages 1–9. IEEE, 2016.

[9] D Raikov. On the decomposition of poisson laws. In *Dokl. Akad. Nauk SSSR*, volume 14, pages 9–12, 1937.

[10] Srikanth Sundaresan, Nick Feamster, and Renata Teixeira. Measuring the performance of user traffic in home wireless networks. In *International Conference on Passive and Active Network Measurement*, pages 305–317. Springer, 2015.

[11] Srikanth Sundaresan, Nick Feamster, and Renata Teixeira. Home network or access link? locating last-mile downstream throughput bottlenecks. In *International Conference on Passive and Active Network Measurement*, pages 111–123. Springer, 2016.

[12] Ilias Syrigos, Stratos Keranidis, Thanasis Korakis, and Constantine Dovrolis. Enabling wireless lan troubleshooting. In *International Conference on Passive and Active Network Measurement*, pages 318–331. Springer, 2015.